

The Expert System for Toxicity Prediction of Chemicals Based on Structure-Activity Relationship

by Masahiro Nakadate,* Makoto Hayashi,* Toshio Sofuni,* Eiichi Kamata,* Yoshitaka Aida,* Takaharu Osada,[†] Tamotsu Ishibe,[†] Yukiko Sakamura,[†] and Motoi Ishidate, Jr.*

The prediction systems of chemical toxicity has been developed by means of structure-activity relationship based on the computerized fact database (BL-DB). Numbers and ratio of elements, side chains, bonding, position, and microenvironment of side chains were used as structural factors of the chemical for the prediction. Such information was obtained from the BL-DB database by Wiswesser line-formula chemical notation. In the present study, the Salmonella/microsome assay was chosen as indicative of the target toxicity of chemicals. A set of chemicals specified with mutagenicity data was retrieved, and necessary information was extracted and transferred to the working file. Rules of the relations between characteristics of chemical structure and the assay result are extracted as parameters for rules by experts on the rearranged data set. These were analyzed statistically by the discriminant analysis and the prediction with the rules were evaluated by the elimination method. Eight kinds of rules to predict Salmonella/microsome assay were constructed, and currently results of the assay on aliphatic and heterocyclic compounds can be predicted as accurately as +90%.

Introduction

To develop new chemicals, a large number of candidates were synthesized and selected according to their characteristics. Not only the usefulness of the chemicals but also the potential of hazard to human health should be considered before the chemicals are introduced to the market. If the toxicity of a chemical can be predicted without any experimentation, a lot of resources, time, and costs are minimized.

Structure-activity relationship studies can also facilitate the identification of hazardous chemicals in our environment because these studies can set the priority for chemicals to be tested by toxicological assays. Structure-activity relationships might also play an important role in elucidating the mechanisms of mutagenesis and carcinogenesis chemically.

There are few computer prediction systems of chemical toxicity available, although it would be desirable to develop such an expert system. The reasons why it is difficult to develop such a system are *a*) difficulty understanding the mechanisms of expres-

sion of chemical toxicity to make prediction rules and *b*) deficiency of the fact database on chemical toxicity. The aim of this study was to develop an expert system to predict the toxicity of chemicals using the fact database (BL-DB) that we have constructed at National Institute of Hygienic Sciences combined with the knowledge of toxicologists.

Architecture of the Expert System

The expert system consists of the fact database system including BL-DB, the data modification module, the rule-making support module, and the expert knowledge-base module (Fig. 1).

Fact Database System

The fact database system consists of the BL-DB as a core database, a data collecting module, a data search module, and a data downloading module. The database stores the data on mutagenicity, carcinogenicity, and teratogenicity of chemical substances. The major part of the database at the present time is information for mutagenicity. The database can expand for other kinds of toxicological data without any alterations of the system. The database stores the toxicological data not only on a single chemical but also combination of chemicals. The BL-DB has three main files: a substance file, a test data file, and a bibliographic file.

*Biological Safety Research Center, National Institute of Hygienic Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158, Japan.

[†]Japan Institute of Synthetic Technology, Hanzomon-Heim, 1-8, Koji-machi, Chiyoda-ku, Tokyo 102, Japan.

Address reprint requests to M. Nakadate, Biological Safety Research Center, National Institute of Hygienic Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158, Japan.

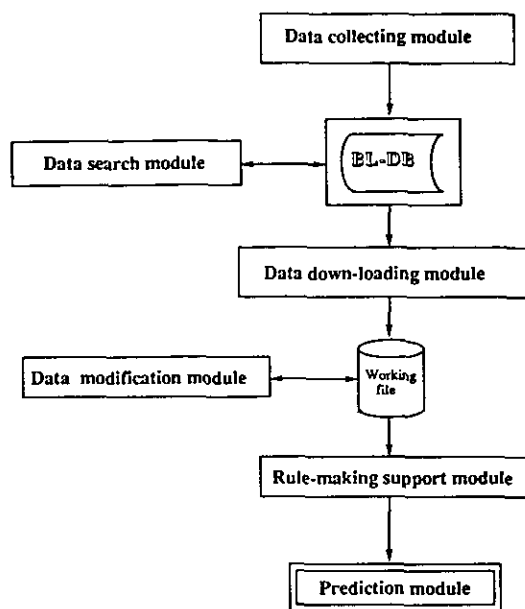


FIGURE 1. Architecture of the expert system.

Formats of data are identical for all toxicological tests, therefore users can access the data with high performance. Many data fields are defined as searchable key fields, therefore users can retrieve target data with many search methods. The structure of chemicals is expressed by Wiswesser line-formula chemical notation (WLN) for the study of structure-activity relation. Sets of data retrieved can be downloaded to the working file for modification and analysis.

Data Modification Module

This module extracts the data to be used for the analysis, such as chemical structure in WLN, molecular formula, and assay results. Also, this module creates substructural fragments in WLN used for descriptions of target chemicals, molecular weight and number of atoms. These modified data are downloaded into the working file to serve them as source data for the rule-making support module. Users can edit data automatically and/or interactively. The edited data can be printed out or output to a file.

Rule-Making Support Module

Using the modified data, this module analyzes and supports to create the rules for predicting toxicity against given chemicals. The Knowledge Acquisition and Utilization System (KAUS), which had been developed by Ohsuga and co-workers (1,2) at Tokyo University, was introduced to support the knowledge-base system. This module has the following functions: *a*) to analyze and classify data according to the chemical structures; *b*) to rearrange and sort WLN and to make the list of KWIC (key word in context) of WLN; *c*) to treat the data statistically, especially by the discriminant analysis; and *d*) to print out the results of analysis to be based on consideration of toxicity prediction rules.

Prediction Module

Using the rules created by the aid of the rule-making support module, the known toxicity of given chemicals is predicted through the prediction module. The results of prediction are printed out and evaluated compared with reported results of toxicity. According to the predictability, these results are used for the feedback to the rule-making support module.

Factors Required for Predicting Chemical Toxicity

Chemicals were classified into 4 major groups and 12 subgroups to make rules. Figure 2 shows the hierarchical groups of structures of chemicals used in the present system.

The factors to be used for predicting chemical toxicity are as follows. Structural factors of chemicals are numbers and ratio of elements, side chains, and bonding (e.g., single, double, triple, position, and microenvironment of side chains). Such information was obtained from the BL-DB database by WLN. Here we chose the mutagenicity by Salmonella/microsome assay as a target toxicity of chemicals.

Strategy to Make Rules for Predicting Chemical Toxicity

A set of chemicals specified with mutagenicity test data was retrieved from BL-DB database. CAS number, chemical name, molecular formula, molecular weight, WLN, result of mutation assay, and other necessary data were downloaded to the working file. Chemicals were classified by their structure. The WLN/KWIC list was printed for each chemical group, which shows positive or negative result in the mutagenicity assay. With the aid of experts, rules of relation between characteristics of chemical structure and the assay result were extracted as parameters for rules. The prediction with the rules was evaluated by the elimination method. These steps were repeated to achieve the most suitable rule for predicting chemical toxicity.

Expert System for Predicting Chemical Toxicity

The flow diagram of the expert system is shown in Figure 3. When a user wants to know the mutagenicity data of chemical

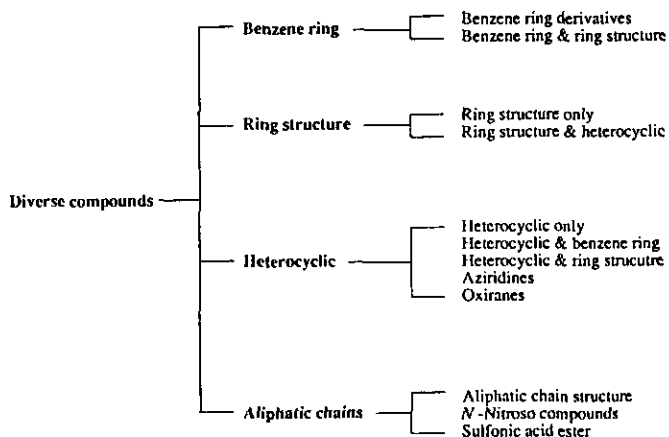


FIGURE 2. Classification of chemicals.

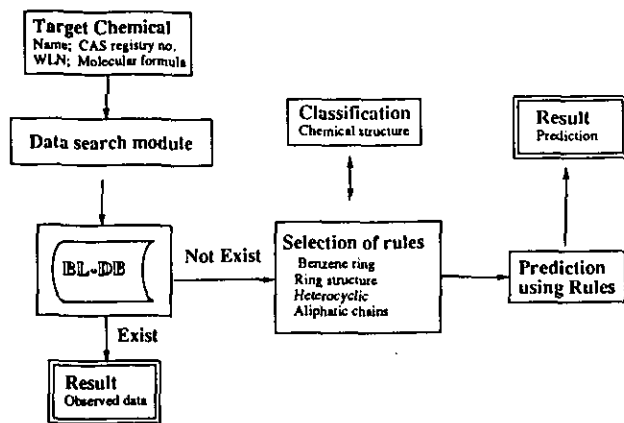


FIGURE 3. The flow diagram of the expert system for prediction.

X, the data are searched in the BL-DB system. If the actual data can be retrieved, the system reports the actual results of assays. On the other hand, if there are no data about X on mutagenicity, the expert system will be used. The structure of X is analyzed and classified by WLN input; thereafter the prediction rules are selected according to the chemical structure. Using the rules selected, the result of the mutation assay is predicted and reported.

In the present study, we constructed eight kinds of rules to predict results of Salmonella/microsome assay. We predicted mutagenicity of chemicals that had not been used to make the rules. The mutagenicity on Salmonella/microsome assay of aliphatic and heterocyclic compounds can be predicted as ac-

curately as more than 90 to 95% at present. We are trying to make new rules including another knowledge base to increase the predictivity using a weight for each parameter.

Limitations

The data from which the rules were constructed might not be sufficient in number; therefore, to increase the accuracy of prediction, an expanding of BL-DB fact database is desired. For an expanding of the database, not only quantity but also quality of data is required. Also an increment of parameters, e.g., partition coefficient of each chemical, and other additional thermochemical properties, can be useful.

To describe the chemical structure, we used WLN in the present study. The connection table of chemical structure might be a better alternative in the next step of prediction system. We are trying to predict the results of the *in vivo* micronucleus assay for a next step.

The final goal of this system is to predict the hazard to human health of a chemical quantitatively not only based on the chemical structure but also combining results of toxicological assay using experimental animals.

This study was supported by funds from Science and Technology Agency of Japan.

REFERENCES

1. Ohsuga, S. Toward intelligent CAD systems. *Computer Aided Design* 21: 315-337 (1989).
2. Yamauchi, H., and Ohsuga, S. Loose coupling of KAUS with existing RDBMSs. *Data Knowledge Eng.* 5: 227-251 (1990).